BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 8: e133172
doi: 10.3897/biss.8.133172

OPEN ACCESS

Conference Abstract

DiSSCover the Potential of FAIR Digital Object Annotations and How You Can Use Them!

Tom Dijkema[‡,§], Sam Leeflang[‡,§]

‡ Naturalis Biodiversity Center, Leiden, Netherlands
§ Distributed System of Scientific Collections - DiSSCo, Leiden, Netherlands

Corresponding author: Tom Dijkema (dijkematom@gmail.com), Sam Leeflang (sam.leeflang@naturalis.nl)

Received: 26 Jul 2024 | Published: 07 Aug 2024

Citation: Dijkema T, Leeflang S (2024) DiSSCover the Potential of FAIR Digital Object Annotations and How You Can Use Them! Biodiversity Information Science and Standards 8: e133172.
https://doi.org/10.3897/biss.8.133172

Abstract

The infrastructure for the Distributed System of Scientific Collections (DiSSCo) is in full development. Work within the DiSSCo Transition Project has been focused on building infrastructure, creating data models, and setting up Application Programming Interfaces (APIs) (Koureas et al. 2024). In the past years, DiSSCo has presented this work at different Biodiversity Information Standards (TDWG) conferences (Leeflang and Addink 2023, Leeflang et al. 2022, Addink et al. 2021). In this year's session, we would like to focus on the human-facing application: DiSSCover.

DiSSCover is the graphical user interface through which users can interact with Findable, Accessible, Interoperable and Reusable (FAIR) Digital Objects (FDOs), facilitating the curation and enhancement of specimen data (Islam 2024). Development started in 2022 and is ongoing. The interface acts as a gateway into the DiSSCo infrastructure, providing access to digital specimens and media. Extracted from the core DiSSCo API, the data is converted into an easily readable format and made discoverable through a diverse set of filters. DiSSCover's main focus is to allow users to make annotations upon the data.

Through the concept of annotations, we connect expert and machine-generated information, to create extended digital specimens (Hardisty et al. 2022), e.g, by creating linkages to other infrastructures, correcting or adding new information, or by triggering machine annotation services. Machine annotation services are automated scalable tools

© Dijkema T, Leeflang S. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

that run in the background and automatically curate and extend the specimen (Addink et al. 2023). Human users will remain important, as all annotations made by machine annotation services can be reviewed by a trusted person. Annotations are Fair Digital Objects and target a specific part of a specimen, be it a data fragment or an associated media file.

At the heart of DiSSCover lies the Open Digital Specimen data specification (Leeflang and Addink 2023). It tries to harmonise multiple data standards into one generic specification based on the new Global Biodiversity Information Facility (GBIF) Unified Model (Robertson et al. 2022). The data is stored as JavaScript Object Notation (JSON) based on JSON Schemas (Anonymous 2024). Annotations are linked to specific data attributes using a JSON-path as the identifier. Data attributes can be individual terms, collections of terms called classes, or the whole object. This creates a flexible but complex data structure, the basis for which we used the World Wide Web Consortium (W3C) web annotation data model (Sanderson et al. 2017).

The W3C Web annotation data model contains two main components: the target and body. The target specifies which data attribute the annotation is made on, for example, the term: 'ods:specimenName'. This is a local term within the open Digital Specimen namespace (ods), which holds the accepted name of the digital specimen. The annotation body holds the value(s) that are appended to the digital specimen, and differ based upon the annotation motivation. DiSSCo recognises five different annotation motivations: addition, modification, comment, assessment and deletion, each of which has its own unique function. This creates a flexible structure that should be able to handle any information the user wants to add to the object. The challenge of DiSSCover is to preserve the complex structure of annotations, whilst making it convenient for users to work with.

The session will provide a look at the different kinds of annotations and their use from a practical perspective. A demonstration of DiSSCover will show how users can create annotations, providing knowledge about the process that will give shape to DiSSCo's main goal of enriching natural history data.

Keywords

user interface, data curation, digital specimen, Digtial Extended Specimen, Open Digital Specimen, DiSSCo, DiSSCo Transition Project, software demonstration

Presenting author

Tom Dijkema

Presented at

SPNHC-TDWG 2024

Funding program

Work on DiSSCover has been funded by DiSSCo Prepare, Grant agreement ID: 871043 (https://doi.org/10.3030/871043); and DiSSCo Transition: Grant agreement ID: 101130121.

Hosting institution

Naturalis Biodiversity Center

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Addink W, Islam S, Alonso J (2021) DiSSCo e-Services to Serve Global Community Needs. Biodiversity Information Science and Standards 5 https://doi.org/10.3897/biss.5.73903
- Addink W, Leeflang S, Islam S (2023) A Simple Recipe for Cooking your AI-assisted Dish to Serve it in the International Digital Specimen Architecture. Biodiversity Information Science and Standards 7 https://doi.org/10.3897/biss.7.112678
- Anonymous (2024) https://schemas.dissco.tech/. Accessed on: 2024-7-11.
- Hardisty AR, Ellwood ER, Nelson G, Zimkus B, Buschbom J, Addink W, Rabeler RK, Bates J, Bentley A, Fortes JAB, Hansen S, Macklin JA, Mast AR, Miller JT, Monfils AK, Paul DL, Wallis E, Webster M (2022) Digital Extended Specimens: Enabling an Extensible Network of Biodiversity Data Records as Integrated Digital Objects on the Internet. BioScience 72 (10): 978-987. https://doi.org/10.1093/biosci/biac060
- Islam S (2024) The data model behind DiSSCo's Annotation service. https://dissco.tech/2024/01/14/the-data-model-behind-disscos-annotation-service/. Accessed on: 2024-7-11.
- Koureas D, Livermore L, Addink W, Alonso E, Alonso J, Casino A, Dusoulier F, Ferreira V, Grieb J, Groom Q, Islam S, Kõljalg U, Lymer G, Marhold K, Paleco C, Pijls S, Scory S, Scott B, Weiland C, Worley K (2024) DiSSCo Transition Abridged Grant Proposal. Research Ideas and Outcomes 10 https://doi.org/10.3897/rio.10.e118241
- Leeflang S, Addink W, Theocharides S (2022) Human and Machine Working Together towards High Quality Specimen Data: Annotation and Curation of the Digital Specimen. Biodiversity Information Science and Standards 6 https://doi.org/10.3897/biss.6.90987
- Leeflang S, Addink W (2023) Harmonised Data is Actionable Data: DiSSCo's solution to data mapping. Biodiversity Information Science and Standards 7 https://doi.org/10.3897/biss.7.112137
- Robertson T, Wieczorek J, Raymond M (2022) Diversifying the GBIF Data Model. Biodiversity Information Science and Standards 6 https://doi.org/10.3897/biss.6.94420
- Sanderson R, Ciccarese P, Young B (2017) Web Annotation Data Model. https://www.w3.org/TR/annotation-model/. Accessed on: 2024-7-11.